Resistant Modelling of Income Distributions and Inequality Measures

RONCHETTI, Elvezio, VICTORIA-FESER, Maria-Pia

Abstract

We review the use and the interpretation of some robustness concepts and techniques in some economic applications. We focus on estimation techniques in income distribution analysis and we discuss the reliability of inequality measures.

Reference

Resistant Modelling of Income Distributions and Inequality Measures

by

Elvezio Ronchetti

Faculty of Economic and Social Sciences
University of Geneva
CH-1211 Geneva, Switzerland

and

Maria-Pia Victoria-Feser

Department of Statistics
London School of Economics and Political Science
Houghton Street
London WC2A 2AE, UK.

Last submitted version for
The Practice of Data Analysis: Essays in Honor of John W. Tukey,
D. R. Brillinger, L. T. Fernholz and S. Morgenthaler (EDs.),

March 1996
Revised June 1996
Abstract

We review the use and the interpretation of some robustness concepts and techniques in some economic applications. We focus on estimation techniques in income distribution analysis and we discuss the reliability of inequality measures.
1 Introduction

Robustness has been one of recurrent element of John W. Tukey’s multiple, diverse, and influential contributions to statistics. His 1960 paper showed the damaging effects of small deviations from the model and was one of the building blocks of modern robust statistics. Robustness ideas played a crucial role in the development of his exploratory data analysis (Tukey 1977) and appeared clearly in his influential 1962 paper and even earlier toward the end of World War II in the work of the Princeton University Fire Control Research Group.

It is therefore fitting on this occasion to discuss some recent developments of robust concepts and techniques. Our goal is not to provide a complete overview, but just to show the usefulness of some basic robustness concepts in some economic applications. A good general discussion of the current status of robustness is given in Huber (1995). Some applications of robustness to econometric models can be found in Koenker (1982), Krasker and Welsch (1985), Peracchi (1990) and Krishnakumar and Ronchetti (1996). Here we focus on models for income distribution analysis (Section 2) and on the derived inequality measures (Section 3).

2 Modelling Income Distributions

Models for income distributions and derived indicators play a central role in the field of welfare economics and inference based on them can strongly affect the economic policies. Since Pareto (1896) fundamental contribution, many classes of models have been derived from either mathematical or economic theory to fit personal income data; for a review see for instance McDonald (1984) and Victoria-Feser (1993). These models are only approximations to the reality, and many studies on income data come to the conclusion that the model does not fit the data well. Common sources of deviations from assumed models include outliers, grouping effects, and other misspecifications of the model; cf. Van Praag, Hagenaars, and Van Eck (1983).

Anomalous observations (outliers) can appear as legitimate high incomes or as recording and definition errors. A not too atypical example of the latter is the week-month confusion when data are supposedly collected on weekly income, but some respondents actually report monthly income. Though high incomes are clearly important from the point of view of the fiscal authority, one can argue that outliers (of both types) should have a limited influence when the goal of the study is to derive inequality measures or other economic indicators that should reflect the economic and social situation of a given
country as a whole. This is even more important when these indicators are fed (automatically) as parameters in more complex econometric models which are used to compare different countries possibly over time.

Grouping effects (due for instance to a shift of some observations from one class to another) appear naturally in income data because the data are typically available only in the form of frequencies per class.

Finally, income data are typically incomplete. There might be truncation problems when too small or too large incomes are not included in the sample, as well as censoring problems when too large or too small incomes are replaced by predefined "minimum" or "maximum" values. A prime example of this is the phenomenon commonly known in the US as "top-coding": many of the higher observations are in effect censored in that lower values are recorded. This is done not only for confidentiality reasons but also for convenience: for a long time the world’s major panel study on income - the University of Michigan’s Panel Study on Income Dynamics (PSID) - suffered from this top-coding problem because incomes exceeding 5-digits could not be stored electronically. In other instances, negative incomes are systematically truncated or put to the "0" default: individuals not being taxed because their income is too low are not included in the sample and large incomes are systematically not recorded for confidentiality reasons. These truncation problems induce systematic biases in the income distribution analysis as shown by Fichtenbaum and Shahidi (1988) with american data.

Given the structure of the data and the possible "unusual features" such as extreme incomes, missing incomes, censored incomes, aggregated data, etc., the models that are supposed to describe them can be only though of approximations to the reality. Robust techniques appear therefore as natural candidates to try to partially alleviate the problems associated with income data. In this section we present some elements of the development of such techniques to the problem of the analysis of income distribution. More details can be found in Victoria-Feser (1993,1996) and Victoria-Feser and Ronchetti (1994,1996).

2.1 Continuous and complete data

Robust estimators for general parametric models when the data are non grouped and complete are given in Hampel, Ronchetti, Rousseeuw, and Stahel (1986). The latter are called OBRE (optimal B-robust estimators) and are actually M-estimators with minimum asymptotic covariance matrix and a bound \( c \) on their influence function \( (IF) \); cf. Hampel (1968,1974). They are suitable candidates for the parameters of income models such as the Gamma distribution which depends on two parameters and the Dagum type
I (Dagum 1980) which depends on three parameters. The first parameter is usually for the scale, the second is for the shape of the distribution and a third parameter is sometimes added to model the thickness of the right tail. As an example, we try to fit one data set to the income distributions mentioned above using the classical maximum likelihood estimator (MLE) and the standardized OBRE.

Our data set is based upon a subsample \((n = 746)\) of a standard data set of disposable income in the UK, 1979, where the income receiver is the household in receipt of social benefits (see Department of Social Security 1992). A Gamma distribution is first fitted. The histogram of the data and the estimated densities (MLE and OBRE, \(c = 2\)) are presented in Figure 1.

We can see that the MLE based on the Gamma model provides a very poor fit, whereas the OBRE captures the bulk of the data in the center of the distribution. One could argue that two parameters are not sufficient to describe this dataset, and that a third one is needed. The question is then: is it worthwhile to add a parameter to model a few extreme values and are we really safe with one more parameter?

To give an element of answer to this problem, we choose to fit a Dagum type I distribution to the same data set. The histogram of the data and the estimated densities (MLE and OBRE, \(c = 2\)) are presented in Figure 2.

We can see that although we have an extra parameter for the thickness of the right tail (to model extreme values), the MLE based on three parameters provides a worse fit than the robust estimator based on two parameters.

### 2.2 Truncated data

To tackle the problem of truncated data, Victoria-Feser (1993) proposes to use either the truncated distribution or a generalization of the EM algorithm (Dempster, Laird, and Rubin 1977) for M-estimators (EMM algorithm). The two estimators are in general different (they are equivalent for the MLE), but their difference is small and due to the assumptions on the underlying model. Contrary to the truncated distribution approach, the EMM algorithm assumes that the underlying model is the complete one. This leads to a difference in the way the weights associated to each observation are computed; see Victoria-Feser (1993).

To show the properties of the OBRE with truncated data, we present a simulation study. Table ?? gives the bias and the MSE for the MLE and the OBRE (EMM) in the case of contaminated and truncated data (below a minimum value), that where generated by a Gamma distribution with scale parameter \(\lambda = 1\) and shape parameter \(\alpha = 3\). The contamination consists on 1% of randomly chosen data ten times their value.
Figure 1: Histogram of the UK data and the estimated Gamma densities (MLE and OBRE, $c = 2$)
Figure 2: Histogram of the UK data and the estimated Dagum type I densities (MLE and OBRE, $c = 2$)
The behaviour of the MLE is not satisfactory, whereas the OBRE has a small bias. In fact, the MLE is even more biased than in the case of complete data (see Victoria-Feser and Ronchetti 1994). This is not surprising, for truncating means incomplete information and robust estimators are constructed to deal with approximate models.

### 2.3 Grouped data

When data are available only in the form of frequencies per class, the OBRE of Hampel et al. (1986) cannot be used as such. The problem can be thought as a discrete data problem with a continuous underlying model. In addition to the usual effects of contamination, we have in this situation grouping effects where some observations may be shifted from one class to another because of rounding errors or class definition. A general class of estimators (Minimum Power Divergence Estimators or MPE) for the parameters of the underlying model based on grouped data was defined by Cressie and Read (1984). It includes the MLE, the minimum Hellinger distance estimator (MHDE), the estimator based on Pearson $\chi^2$ etc. Though their IF is bounded, deviations from the underlying model can cause a large bias especially in classes with low probabilities.

Victoria-Feser and Ronchetti (1996) propose a more general class of estimators containing the MPE, called MGP, which can be seen as M-estimators for grouped data. An optimal estimator (OBRE) that minimizes the asymptotic covariance matrix under a bound on its IF is derived, and it is shown that for a small efficiency loss, it is more robust than its classical (MPE) counterpart.

We illustrate the performance of MGP estimators with a small simulation from a Pareto distribution (a classical one-parameter income distribution model). The data are contaminated by taking $\varepsilon$ proportion of randomly chosen observation and multiplying them by 10, and then grouped in 22 classes.

The MPE and OBRE (for the MLE and MHDE) are computed and their Bias and MSE reported in Table ?? (the standard errors for the values of the bias are less than 0.02).
We can see that although the *IF* of the MLE is bounded, when the underlying model is contaminated, the MLE has a large bias. On the other hand, with the corresponding robust estimator we can see that this bias and the overall MSE are considerably smaller. The MHDE has better robustness properties than the MLE, but it can be improved by using the corresponding robust version which has the best bias and MSE overall in the example.

The robust procedure is also applied to the UK data set presented above. The data are grouped in 58 equal size classes. The first class is extended to 0 and the last class to ∞. The Gamma distribution is chosen as candidate to model the data and the MLE, MHDE and robust MHDE are computed. The histogram of the data and the estimated densities are presented in Figure 3.

While the MLE tries to accommodate the tails of the distribution, it misses the description of the bulk of the data in the center. The MHDE improves the fit of the majority of the data but it is its robust version which gives the best fit. The latter has an efficiency of 95% compared to the classical MPE.

Victoria-Feser and Ronchetti (1996) show that the OBRE for grouped data has a bounded local shift sensitivity and therefore it controls the effects of grouping errors. This is not the case with classical MPE.

## 3 Inequality Measures

Inequality measures and other related tools of income distribution analysis are often used to summarize information about income distributions. They play an important part in political debates about economic and social trends and in welfare economics.

At a theoretical level, inequality measures have been derived by requiring a number of essential properties. The most important (and most accepted) property is the principle of transfers (Dalton 1920), which states that the transfer of an arbitrary positive amount of income from a poorer income receiver to a richer one (such that the mean of the distribution is preserved) should increase the value of the inequality measure.

On the other hand, inequality measures are estimated from income data. Therefore, it is important to understand the relationship between the eco-
Figure 3: Histogram of the grouped UK data (58 equal size classes) and the estimated Gamma densities (MLE, MHDE and robust MHDE)
onomic properties that the inequality measure should fulfill and the statistical properties of the corresponding estimator. In this section we summarize some of the results obtained in Victoria-Feser (1993) and Cowell and Victoria-Feser (1996) on the relationship between economic and robustness properties of inequality measures.

The aim is to find a simple and convenient way to check if an inequality measure is resistant against data contamination and draw conclusion about general classes of inequality indices. It can be shown that the principle of transfers alone does not imply necessarily non resistance of an inequality measure. However, if another property is added, like the scale independence or decomposability of the index (see Cowell and Victoria-Feser 1996) which restricts the class but still encompasses most of the widely used inequality indices, or if a more realistic specification of the estimation problem is considered (like the mean income has to be estimated rather than being specified a priori), then it can be shown that any inequality index satisfying these properties is not robust in that its \( IF \) is unbounded. In particular this is the case for the Kolm index (Kolm 1976a and Kolm 1976b), the generalized entropy family (Cowell 1980) which includes the Theil indices (Theil 1967) and the Gini index (Gini 1910).

To show the effect of data contamination on the estimator of the inequality, 100 samples of 200 observations from a Pareto distribution are generated and contaminated by multiplying by 10 a proportion of randomly chosen observation. Three indices are computed, the Gini index, the Theil index and a coefficient of variation type index (CV) and the average values of these indices are given in Table 4. (The second row represents the true values of the indices and SD stands for standard deviation.)

<table>
<thead>
<tr>
<th>Contamination</th>
<th>Gini SD 0.2</th>
<th>Theil SD 0.0945</th>
<th>CV SD 0.1667</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.1974 0.002</td>
<td>0.0908 0.003</td>
<td>0.1545 0.02</td>
</tr>
<tr>
<td>1%</td>
<td>0.2577 0.002</td>
<td>0.2068 0.005</td>
<td>0.4974 0.02</td>
</tr>
<tr>
<td>2%</td>
<td>0.2845 0.002</td>
<td>0.2584 0.006</td>
<td>0.6466 0.03</td>
</tr>
<tr>
<td>3%</td>
<td>0.3319 0.002</td>
<td>0.3485 0.007</td>
<td>0.8993 0.04</td>
</tr>
<tr>
<td>4%</td>
<td>0.3542 0.002</td>
<td>0.3920 0.007</td>
<td>1.0180 0.04</td>
</tr>
<tr>
<td>5%</td>
<td>0.3891 0.002</td>
<td>0.4500 0.007</td>
<td>1.1150 0.04</td>
</tr>
</tbody>
</table>

We can see that the three indices are very sensitive to small contamination in the data. Cowell and Victoria-Feser (1996) contains a more complete study involving other income distribution models.

To avoid the effect of contamination on the estimated indices, one way is to estimate inequality from robust estimates of the parameters of income distribution models. Cowell and Victoria-Feser (1996) show for example that
for the Entropy family indices, the $IF$ of the members of this class is proportional to the $IF$ of the estimators of the parameters of the underlying income model. Therefore, if robust estimators for the parameters of the income model are used, then the resulting inequality measure will also be robust. These robust estimators of inequality convey a lot of information in that they provide a check against classical estimates; where discrepancies between the results emerge and are attributable to small deviations from the assumed model, this information should be taken into account in drawing conclusions about the "true" picture of inequality.

For example, using the UK dataset, the three inequality indices presented above are estimated through the MLE and OBRE of the Gamma distribution and compared to direct (empirical) estimates on the whole sample and on a truncated sample in which only 2% of the top incomes have been removed. This is to show to what extent the inequality measures considered here are sensitive to extreme values. The results are presented in Table ??.

<table>
<thead>
<tr>
<th></th>
<th>Gini</th>
<th>Theil</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.1375</td>
<td>0.0299</td>
<td>0.0302</td>
</tr>
<tr>
<td>OBRE ($c = 2$)</td>
<td>0.0983</td>
<td>0.0152</td>
<td>0.0153</td>
</tr>
<tr>
<td>Direct estimates on whole sample</td>
<td>0.1287</td>
<td>0.0320</td>
<td>0.0358</td>
</tr>
<tr>
<td>Direct estimates on truncated sample</td>
<td>0.112</td>
<td>0.0222</td>
<td>0.0228</td>
</tr>
</tbody>
</table>

They reveal two interesting points. First we can see that the OBRE produces a distribution that exhibits uniformly lower inequality than that produced by the MLE (even by a factor 2!). Secondly, the same phenomenon can be observed by comparing the two direct estimates; the upper truncation of the sample has an impact upon inequality that is similar to switching from the MLE to the OBRE.

## 4 Conclusions

We showed that robust techniques can play a useful role in income distribution analysis by providing more reliable fits and more stable inequality measures. Further research developments are needed to provide robust inference and model selection for grouped data. Moreover, the use of robust procedures on a large data base like the Luxembourg Income Study which makes data available for a wide range of countries, will provide guidance on the calibration and usefulness of the robust analysis.
References


